Cardiff University Press
Gwasg Prifysgol Caerdydd

JCaDS
JOURNAL OF CORPORA AND DISCOURSE STUDIES 2018, 2:14-39
E-ISSN 2515-0251

'SO IS YOUR MOM AS CUTE AS YOU?': EXAMINING PATTERNS OF LANGUAGE USE BY ONLINE SEXUAL GROOMERS

NURIA LORENZO-DUS &
ANINA KINZEL
SWANSEA UNIVERSITY

CITATION

Lorenzo-Dus, N. & Kinzel, A. (2019). 'So is your mom as cute as you?': examining patterns of language use by online sexual groomers. Journal of Corpora and Discourse Studies, 2(1):1–30

ABSTRACT

Linguistic research into online grooming is scarce despite both the communicative essence of this form of online child sexual abuse and a substantial body of literature into it across other Social Sciences. Most of this literature has examined small data sets via qualitative methods, primarily Thematic Analysis. This study evaluates the contribution that a Corpus Assisted Discourse Studies (CADS) approach can make to this body of literature, with a focus on online groomers' language. The corpus, extracted from the Perverted Justice Foundation archive, consists of c. 3.3 million words produced by >600 convicted child sexual offenders interacting online with adult decoys whom they believed to be children. Lexical dispersion (DP_{Norm}), collocation and concordance analyses were conducted. The corpus was also run through the software LIWC (Linguistic Inquiry and Word Count), which is the only other software-assisted methodology that has been adopted to examine Perverted Justice data. Our analysis shows that LIWC may not be the most efficient software to analyse online groomers' language due to a lack of general language comparison scores, the non-transparency of some of its analytic variables and a focus on de-contextualised words. Comparatively, CADS methods can shed light upon online groomers' strategic use of language. They can also reveal the complex and nuanced ways in which discourse features such as sexual explicitness/implicitness and interpersonal (in)directness operate alongside these strategies.

KEYWORDS

Online sexual groomer language; discourse; corpus; software; Linguistic Inquiry Word Count (LIWC); sexual desensitisation, isolation, risk assessment, deceptive trust

CONTACT

Nuria Lorenzo-Dus, Department of Applied Linguistics, College of Arts and Humanities, Swansea University, Wales, UK. n.lorenzo-dus@swansea.ac.uk

ORCID

0000-0002-6211-7939

DOI

https://doi.org/10.18573/ jcads.31

ISSUE DOI

https://doi.org/10.18573/jcads.v2i1

Manuscript accepted 2019-03-08

'So is your mom as cute as you?': examining patterns of language use by online sexual groomers

Nuria Lorenzo-Dus
Swansea University

Anina Kinzel
Swansea University

1. Introduction

Sexual grooming of children online is a legally punishable form of Child Sexual Abuse (henceforth, CSA), irrespective of whether it leads to off-line (physical) contact between the groomer and the child.[1] Children who suffer online grooming/CSA are known to experience self-blame, depression and low-self-esteem; they may also subsequently self-harm and develop behavioural problems (Hamilton-Giachritsis *et al.,* 2017). The impact of their communicative involvement online with a sexual predator can, in short, be as damaging as that ensuing from physical CSA (National Society for the Prevention of Cruelty to Children [NSPCC], 2017).

The seriousness of the offence and its increasing prevalence over time (Bentley *et al.,* 2018) may help account for a growing body of academic scholarship into online grooming over the past decade or so. This scholarship has been primarily conducted within the disciplines of Criminology and Psychology. In comparison, and despite a broad understanding that online grooming essentially entails communicative manipulation for the purpose of sexual abuse, its study within Linguistics is significantly underdeveloped. Moreover, it has to date tended to rely on analysis of small data sets, which makes it difficult to identify patterns in online groomers' communicative *modus operandi*. In light of this, the main aim of this article is to explore the potential contribution of a Corpus Assisted Discourse Studies (CADS, henceforth) approach to the study of a large corpus of online grooming interactions ('chat logs') extracted from the publicly available Perverted Justice archive (see Section 2), specifically the language used by online groomers therein. Two studies in Criminology and Psychology (Black *et al.,* 2015; Drouin *et al.,* 2017) have applied software-enabled methods, specifically Linguistic Inquiry and Word Count (LIWC http://liwc.wpengine.com/), to the analysis of a large number of chat logs from Perverted Justice. A subsidiary aim of our article, therefore, is to evaluate to what extent LIWC – as applied in those studies – and CADS methods may complement each other.

The article is structured as follows. In Section 2 we review extant research into the *language* used in online grooming. Section 3 describes our corpus and analytic procedures. In Section 4 we report and discuss our findings, focusing on online groomers' sexual

1 In a UK context, for instance, this is covered by the Serious Offences Act (2003) (https://www.legislation.gov.uk/ukpga/2003/42/contents) and the Serious Crime Act (2015), which made it a criminal offence for an adult to send a sexual message to a child (http://www.legislation.gov.uk/ukpga/2015/9/section/67/enacted).

desensitisation (Section 4.1) and risk assessment (Section 4.2) processes. Section 5 evaluates the potential contribution of our findings to a better understanding how online grooming works communicatively.

2. Online Grooming: A Focus on Language

Sexual grooming of children online is not an easy concept to define – or characterise. An oft-cited definition is that provided by Craven *et al.* (2006, p. 297) for offline grooming, namely '[a] process by which a person prepares a child, significant adults and the environment for the abuse of this child'. Research into online grooming has tended to adapt this definition to the online environment by adding that the process is Internet-facilitated or enabled (see, e.g. Katz, 2013; Kloess *et al.,* 2017; Whittle *et al.*, 2014; Winters *et al.*, 2017; Lorenzo-Dus *et al.,* 2016; Chiang and Grant, 2018). A few studies have explicitly argued that the *digital* environment in which this form of CSA occurs crucially shapes the sexually abusive relationship that online groomers develop with their targets. Unlike offline grooming, for instance, online grooming is not a linear process that goes from an adult befriending a child to engaging him/her in sexual abuse. Instead, online groomers activate a complex 'communicative entrapment network', resorting to overlapping, non-sequential manipulation strategies (Lorenzo-Dus *et al.*, 2016, p. 46; see also, Williams *et al.*, 2013; Van Gijn-Grosvenor and Lamb, 2016; Quayle and Newman, 2017). [2]

These strategies inform a number of 'online grooming models' that have been developed within the fields of Psychology and Criminology using primarily a Thematic Analysis methodology and examining relatively small data sets (see, for example, Egan *et al.*, 2011; Kloess *et al.*, 2017; Williams *et al.*, 2013; Quayle & Newman, 2016; Winters *et al.*, 2017). While labelling its components differently[3], these models broadly encompass online groomers' attempts at developing the child's trust in them (terms used include: relationship forming, rapport building, friendship forming), obtaining sexual gratification from the online interaction with the child (terms used include: sexual content, sexual rapport), testing the child's compliance level (terms used include: risk assessment, assessing and managing risks) and approaching the child for an offline meeting (terms used include: meeting at offender's house, contact request, meeting planning).

The comparatively scarce and late-developing Linguistics literature into online grooming has expectedly, and explicitly, highlighted its communicative essence

2 A few studies do see online grooming as a linear process (O'Connell, 2003; Egan *et al.*, 2011; Black *et al.*, 2015; Kloess *et al.*, 2017).

3 The terminology for the components of these models is wide-ranging, including 'strategies', 'themes', 'processes', 'stages', 'tactics' and 'moves'. In this paper, we use 'processes' and 'strategies' as per the model of online grooming discourse presented in Figure 1. A detailed review of each of these models, as well as other areas of non-linguistic scholarship into online grooming, such as victim and perpetrator typologies and motivations, is beyond the scope of this paper.

(Lorenzo-Dus *et al.,* 2016; Chiang and Grant, 2017). Building upon extant knowledge into online grooming from other disciplines, it has sought to characterise how online groomers use language – that is, their discourse. Figure 1 reproduces the first model of online grooming *discourse* (Lorenzo-Dus *et al.,* 2016). The model is derived from Computer-Mediated Discourse Analysis (Herring, 2013) of data from the Perverted Justice website (http://www.perverted-justice.com/), which is a publicly-available archive of online chat logs between convicted child sexual offenders and adults masquerading as children (labelled 'decoys'). Only the online discourse generated by the offenders was examined.

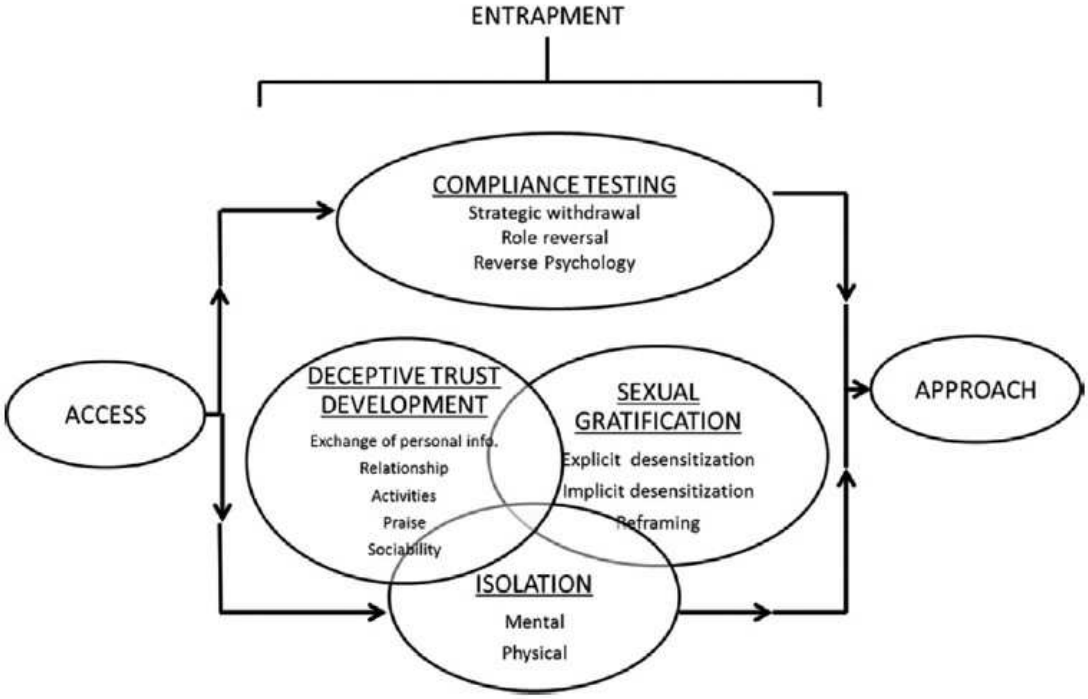

Figure 1: Online Grooming Discourse Model (Lorenzo-Dus et al., 2016:44)

As Figure 1 shows, there are three main phases in online grooming: access, entrapment, and approach. The access and approach phases respectively designate groomers' initiating contact with their target online (an opening turn, such as a greeting) and securing offline access to him/her for the purpose of sexual abuse. The entrapment phase includes four overlapping processes, namely: deceptive trust development, compliance testing, isolation and sexual gratification.

Through the deceptive trust development process, groomers hide their ulterior motive of wanting to engage the target in sexual activities by discursively building a trust-based bond with him/her. This involves using one or more of the following strategies: eliciting and sharing personal information, discussing activities, being sociable through small talk, praising him/her via compliments and discussing relationships with others (family, friends, ex-partners) and the groomer. Compliance testing is a process by

which groomers check whether their target is actually underage and the likelihood of him/her participating in whatever sexual activity groomers propose without being discovered by others. It is realised via three strategies: challenging the target to undertake activities that advance the abusive relationship (reverse psychology), seemingly putting the target in control of the grooming process (strategic withdrawal) and assuming a low-risk attitude from the target in relation to sexual activity (role reversal). The isolation process entails constructing the groomer – target relationship as exclusive and therefore seeking to distance the latter emotionally and physically from others who are close to him/her. Groomers do this through requests to talk to their target privately and alone, making sure he/she is not supervised and instructing him/her to delete any trace of his/her interaction with them (physical isolation). Groomers also seek to separate their target affectively from his/her support network, typically from his/her friends and family (mental isolation). Sexual gratification covers groomers' discursive attempts to involve the target in sexual activities online and, in the case of so-called 'contact-driven' groomers, preparing him/her for sexual interaction offline. This is done by introducing overt visual / textual sexual content (explicit desensitisation). It is also done via more covert methods that rely on textual indirectness/implicitness as well as references to romance rather than sexual behaviour (implicit desensitisation). Sexual gratification can involve reframing, too, whereby groomers discursively present the proposed sexual activity as being beneficial to the target in his/her future life, often by enacting teacher (adult groomer) - pupil (presumed child) discourse roles.

Chiang and Grant's (2017, 2018) work illustrates this nascent interest in the linguistic study of online grooming. Their 2017 study examined Perverted Justice chat logs; their 2018 study analysed chat logs from a single offender interacting with children online. In both studies, the authors used a genre analysis framework to identify different rhetorical moves used by online groomers. Their 2018 study revealed 16 such moves,[4] 14 of which had been identified in Chiang and Grant (2017). These 14 moves are present, under different labels, in other taxonomies of online grooming, thus supporting their identification as being characteristic of online groomer discourse. The two additional moves in Chiang and Grant (2018) were: 'overt persuasion' and 'extortion'. Overt persuasion refers to groomers' discursive attempts to exert power over the victim by, for instance, commanding, threatening, seeking sympathy and material offers. Extortion involves groomers' use of direct and indirect threats, blaming the victim and 'stating "contractual" terms' of their online 'relationship' (Chiang and Grant, 2018, p. 11).

Chiang and Grant (2018) note that these two moves had not been observed in the previous literature, which used Perverted Justice data, thereby calling into question its

4 Greeting, maintaining conversation, meeting planning, reprimanding, sign-off, rapport, assessing likelihood and extent of engagement, assessing criteria fulfilment, assessing and managing risk, assessing role, sexual rapport, initiating sexual topics, maintaining/escalating sexual content, immediate sexual gratification, overt persuasion, extortion

use 'as good data proxy for research into genuine online CSA [Child Sexual Abuse] conversations' (2018, p. 19). Albeit under different labels, use of overt persuasion and extortion has been previously referenced in the online grooming literature. For instance, groomers were found to use threats to ensure compliance (Quayle and Newman, 2017), 'including repetition and entrapment to force the child to talk about sexual topics' (Williams *et al.,* 2013, p. 150) and blackmail to convince the children to engage in sexual activities online (Kloess *et al.,* 2017). Data from Perverted Justice was used in Quayle and Newman (2017) and Williams *et al.,* (2013).

In a research note responding to Chiang and Grant (2017, 2018), Schneevogt, Chiang and Grant used Corpus Linguistics techniques to identify 'examples of overt persuasion or sexual extortion' (2018, p. 98) in 622 Perverted Justice chat logs. The study's methodological procedure, which is only briefly outlined, entailed eliciting a list of 'key terms' (words, bigrams and trigrams) linked to overt persuasion or sexual extortion in the corpus by using both a reference corpus (the Brown corpus) and findings from Chiang and Grant (2018). This led to identifying a smaller and different number of examples of persuasion to those in Chiang and Grant (2018), and none of extortion. Although Schneevogt *et al.* (2018) concluded that Perverted Justice data is not 'truly representative of interactions between child victims and their abusers' (2018, p. 97), they also allowed that the 'data may still be useful for asking some important questions' and stressed the importance of researchers across disciplines being 'aware that in analysing PJ [Perverted Justice] chat logs they are not in fact analysing conversations which involve the abuse of children' (2018, pp. 101-2).

Scholarship into online grooming has by and large examined relatively small data samples. Within Linguistics, Lorenzo-Dus *et al.*'s (2016) study considered 24 chat logs (c. 80,000 words produced by a total of 24 different groomers); Lorenzo-Dus and Izura (2017) analysed all the compliments (n=1,268) in 68 chat logs by the same number of online groomers; Chiang and Grant (2018) examined 20 chat transcripts from one single offender, and Chiang and Grant (2017) analysed seven chat logs from seven online groomers.

There are only three studies – to our knowledge – that have examined online groomer *language* within larger data sets: Black, Wollis, Woodworth and Hancock (2015), Drouin, Boyd, Hancock and James (2017) and Schneevogt *et al.*, (2018). As noted above, the latter study is a brief research note seeking to confirm whether or not Perverted Justice data may differ from the data used in Chiang and Grant (2018) as regards groomer use of overt persuasion and/or sexual extortion. For their part, Black *et al.,* (2015) and Drouin *et al.,* (2017) analysed Perverted Justice chat logs using LIWC.

LIWC was first developed in 1994 within the field of Psychology and has since then been applied across disciplines, ranging from Health (e.g. Park et al., 2012; Wilson and Valstar, 2014; O'Dea et al., 2017) to Terrorism (e.g. Vergani and Bliuc, 2018) Studies. The

software contains pre-defined dictionaries that are aligned to 'variables' that concern psychological metrics (for example, different types of emotions: anger, sadness, happiness and so forth), semantic fields (for instance, sex, family, home) and grammatical categories (for example, verbal tense and word class). Words in the corpus to be examined through LIWC are counted against these variables. The most recent version of the software, LIWC 2015, includes 92 variables. Some variables combine other variables. For example, the variable 'clout', which measures 'expertise' (Pennebaker et al., 2015b, p. 22), was derived from a paper by the LIWC creators (Kacewicz et al., 2014) in which they attempted to identify which language dimensions were connected to a person's place in a social hierarchy. Although the underlying algorithm for this variable has not been disclosed in the LIWC literature or manuals, from the conclusion in Pennebaker et al.'s (2015b) study one may infer that it is in part connected to pronoun use.

LIWC works by counting words within a given variable and comparing them to the overall word count of the data set being examined, assigning a percentage to each variable (Tausczik and Pennebaker, 2010; Pennebaker et al., 2015a). Aggregated variables are calculated on a scale from 0 to 100 - the higher the score the more prevalence of that aggregated variable in the data. LIWC thus performs analyses of textual data sets at the decontextualized, single word level. LIWC semantic and psychological (as well as other types of) variables have been used in research that sought to characterise how certain topics, such as depression (De Choudhury, 2013) and suicide (O'Dea et al., 2017), are talked/written about in large collections of texts and to profile their authors psychologically. Doing so on the basis of automated (software-enabled), decontextualized, single word analysis – as enabled by the LIWC software – disregards the proven link between the textual context in which words are used and their meaning. We share Brookes and McEnery's (2018) view, in relation to the use of topic modelling approaches, that quantitative, largely decontextualized, single word based methodologies 'should not replace close, qualitative analysis of the texts in the data' (2018, p. 17). Instead, an integration of automated and manual methods should be adopted, to be guided by specific research questions.

Returning to LIWC based studies of online grooming, Black et al. (2015) examined 44 online grooming chat logs, dividing each of them into five consecutive segments. They then mapped the results of the LIWC analysis of their data onto online grooming 'stages', taken from O'Connell's (2003) model of online grooming, namely: friendship forming, relationship forming, risk assessment, exclusivity and sexual. For example, risk assessment was aligned to the 'family', 'they' and 'negative emotion – anxious' and 'home' variables in LIWC. The rationale behind such mapping of online grooming stages, variables and decontextualized, single-word usage was not fully specified. Their analysis revealed that some of the online grooming stages tended to occur within certain segments. For example, the risk assessment stage occurred mostly in the first and second segments of the chat logs, while the exclusivity stage was used more in the fourth and fifth segments. The

sexual content stage was introduced as early as the first segment of the chat log. Terms contributing to the friendship and relationship forming stages occurred throughout the chat logs.

For their part, Drouin et al. (2017) analysed 590 online grooming chat logs from Perverted Justice, dividing their data up by supposed decoy gender. The analysis entailed comparing groomer and decoy language using three LIWC variables: 'word count', 'sexual words' and 'clout'. They found that (i) groomers used more sexual words than decoys, even more so in cases of grooming female decoys, (ii) groomers used more words overall than decoys, and (iii) groomers had higher clout scores than decoys. Drouin et al. (2017) also correlated the same three LIWC variables and groomers' sentence outcome regarding jail time. They found a positive correlation between overall words used and length of jail time groomers received but no correlation between a higher sexual word count and a longer jail time.

3. Methodology

The corpus for our study comprises the groomers' contribution within chat logs available from the Perverted Justice Archive. This amounted to over 600 different chat logs, and 3,300,958 word tokens. The majority of the chat logs came from Instant Messenger platforms and took place between 2004 and 2016.

Data extraction proceeded as follows. Firstly, we scraped the chat logs from the Perverted Justice website using a custom Python script, transferred them to a data-frame and saved them to different csv files (one for each chat session within each chat log). A metadata spreadsheet was next created, where further details for each chat log were manually annotated, including the groomer / decoy(s) usernames and the digital platform used for the interaction. Then we used another custom Python script to convert the extracted csv files into XML, which is the format supported by CQPWeb (Hardie, 2012) - the software that we used for analysis. The script also (i) separated each message in the chat log as a single writing turn, distinguishing between user types; (ii) identified emoticons, replacing them with their respective labels, for example :o) → HAPPY-SMILEY; and (iii) deleted any UTF-8 character longer than 4 bytes (as CQPWeb does not support such characters). Finally, and upon realising the high incidence of non-standard spellings (i.e., acronyms, abbreviations, and typographical errors) in the data, we proceeded to normalise it. To do so, we selected approximately one quarter of the whole data and imported it into VARD 2.5.4 (Baron, 2018), identifying and manually standardising all the spelling variations found. We then used these manually normalised variations as training data to standardise the whole corpus automatically.

We next applied the following analytic steps to our corpus:

- We ran it through LIWC 2015, selecting the same variables that were examined in Black *et al.,* (2015) and Drouin *et al.,* (2017).

- We conducted a CADS analysis, namely:
  a. We firstly used #LancsBox 3.03 (Brezina *et al.*, 2015) to calculate lexical dispersion in the corpus, using Gries' (2008, 2010) DP_{Norm} (normalised deviation of proportions) measure. The DP_{Norm} measure 'ranges from 0 to 1, where values around zero mean that the relative frequencies of occurrence of the linguistic expression in the corpus are directly proportional to the sizes of the corpus parts whereas large values close to 1 mean the linguistic expression is distributed very unevenly among the corpus parts' (Gries, 2010, pp. 274-75).
  b. We then used the Dice coefficient metric (CQPWeb [Hardie, 2012]; see also Baker and Levon, 2015; Gabrielatos, 2018) to calculate the collocates of the content words identified in a).
  c. We next manually examined extended concordances of the top 15 collocates in b).

Given space constraints, and our subsidiary aim, in the next section we report the results pertaining to the online grooming strategies that were also the focus of Drouin *et al.*, (2017) and Black *et al.'s,* (2015) work: sexual desensitisation (Section 4.1) and risk assessment (Section 4.2).

4. Results

4.1. Sexual Desensitisation in Online Grooming

Drouin *et al.* (2017) selected the LIWC variables `word count`, `clout` and `sexual content` to examine online groomers' use of language. We did not consider either word count or clout. The former was only relevant to their aim to compare groomer and decoy behaviour, specifically which party was lexically more / less dominant. Our reason for not considering the clout variable was two-fold. Firstly, in LIWC `clout` is an aggregated variable based on a closed source-code algorithm, which means that it is not possible to replicate its performance using different software tools – those provided within a CADS approach in our case. Secondly, use of `clout` as a measure of influence in digital communication is contested (see, for example, Cha *et al.*, 2010; Himelboim *et al.,* 2009; Quercia *et al.*, 2011). Our own view is that a groomer's degree of influence online over his/her target cannot be simply measured through static linguistic features, such as particular pronoun usage. Rather, digital influence is a more complex and dynamic process (Lorenzo-Dus and Di Cristofaro, 2016) – one that, like topic/theme identification in large corpora, operates at a broader contextual level (see Van Dijk, 1980; and Brookes and McEnery (2018) for a critical evaluation of topic modelling approaches).

As for LIWC's sexual content variable, this was relevant for us to consider given both the presence of sexual language in online grooming (as attested by all the literature about this issue) and the need to determine how online groomers embed sexual content into their discourse. As noted in Section 2, the various strategies that online groomers use are

known to exhibit a high degree of overlap and we also know that, as regards attempts at developing trust for deceptive purposes for example, online groomers intersperse sexually and non-sexually oriented compliment topics. Moreover, groomers' discursive work at seeking sexual gratification is known to entail implicit and explicit sexual desensitisation of their targets. All things considered, therefore, we felt it was appropriate to focus on the sexual content variable in LIWC2015 for this study.

Table 1 shows the scores for sexual content terms in our corpus and the corpus used by Drouin *et al.* (2017).

Male (M) grooming		**Female (F) grooming**		**M & F grooming**	
Our study	**Drouin *et al.* (2017)**	**Our study**	**Drouin *et al.* (2017)**	**Our study**	**Drouin *et al.* (2017)**
N= 48	N= 52	N=562	N=583	N=610	N= 590
0.99%	1.03%	1.07%	1.09%	1.07%	--*
3.42%	--*	5.21%	--*	5.21%	4.80%
0.00%	--*		--*	0.00%	0.00%

Table 1: *Sexual content terms. *Not provided in study*

The scores in Table 1 are similar, which is to be expected given the significant overlap between the two data sets on which they are based. The highest frequency of use of sexual terms, for instance, ranges between 5.21% (our corpus) and 4.80% (Drouin *et al.*'s corpus). The scores also show little difference in the use of sexual terms when grooming male and female decoys, the percentage being slightly higher in the case of the female decoys (decoys identified their gender in all the cases examined). These scores may be informative in terms of ascertaining, within each corpus, the proportion of sexual content words vis-à-vis scores for other variables, such as, `family`. Given the overall sexual abuse aim of online grooming, one may be tempted to interpret that the scores in Table 1, whether taken as an average (1.07% - our study) or maximum value (5.21% - our study; 4.80% - Drouin *et al.*'s study), are rather low. However, in the absence of general language use benchmarks for any of the variables in the LIWC dictionaries, LIWC scores do not allow cross-corpora comparisons. As such, it is not feasible to infer from the scores in Table 1 whether they are high or low.

As for the actual words that were responsible for the scores in Table 1, the LIWC2015 dictionary for the sexual content variable comprises a total of 131 words. These include a wide range of terms about sexual matters, ranging from sexual orientation (e.g. `bi-sexual`, `heterosexual`) and sexual organs (e.g. `penis*`, `vagin*`, `womb`) – as well as slang terms for both (`dyke`, `homo`, `dick`, `pussy`) through to sexually transmitted diseases and infections (e.g., `gonorrhoea`, `hiv`, `chlamydia`), sexual violence and assault terms (e.g. `rape`) and sex enhancements (e.g. `viagra`, `vi-`

bratorr). The list also includes a mixed bag of terms linked to sexuality (e.g. libido, lust). It is worth noting that the LIWC2015 dictionary for this variable includes neither euphemistic nor implicit terms for sexual content.

As for the CADS analysis, our corpus contained 336 words that denoted sexual content, all of which were included in the LIWC 2015 'sexual content' variable. However, only 85 of them occurred 10 times or more in our corpus (see Table 2a). Additionally, and as shown in Table 2, there was a series of words with DP_{Norm} scores between 0.6 and 0.7 that had sexual connotation and were not listed in the 'sexual content' variable in LIWC2015.

Rank	Word	DP_{Norm}	Frequency
293	suck	0.63	1545
355	touch	0.63	1177
316	ass	0.68	1419
335	wet	0.70	1278
378	lick	0.70	1086

Table 2: *Frequency and DP_{Norm} scores for sexual content terms in our corpus that were not listed in LIWC 'sexual content' variable*

Given their DP_{Norm} scores and frequencies of use, the words in Table 2 may be said to be reasonably dispersed amongst the chat logs in our corpus. Collocate and concordance analyses of these words confirmed that they all were semantically tied to sexual content. Online groomers used them as part of their sexual desensitisation strategy. Owing to space constraints, we next illustrate and further discuss this finding through two of these words, namely touch and lick.

Figure 2 illustrates the top collocates for touch and lick. The R and L in the graphs denote whether the collocate appears on the left or right of the node word (touch/lick) (dice coefficient values for touch and lick collocates in Figure 2 appear in the Appendix, respectively, Tables 4 and 5).

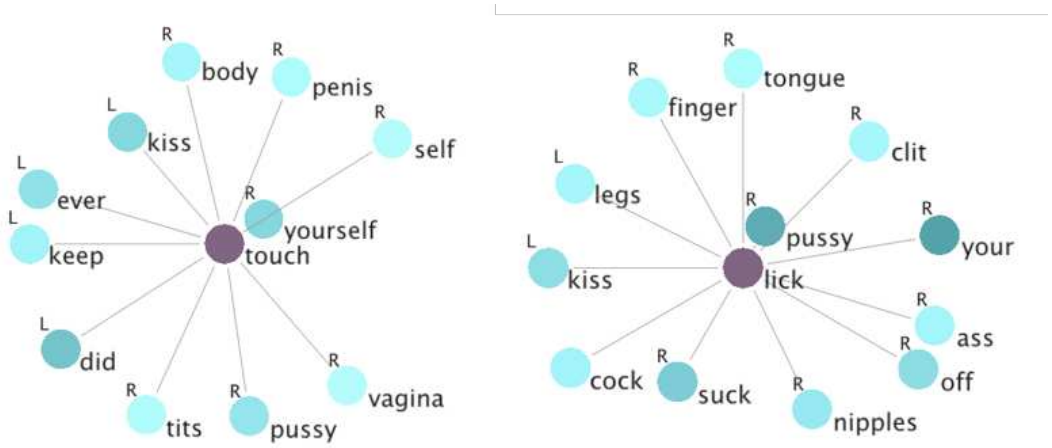

Figure 2: *Collocation graphs for 'touch' and 'lick'*

In the case of touch, four of the 11 collocates were sexual organs (penis, vagina, pussy, tits) and four combined with the word touch to make explicit reference to sexual activity (touch + yourself/self; touch + kiss; touch + body). Of the remaining three, two (touch + ever/did) also served the online grooming strategy of sexual desensitisation (keep + touch was used as part of the prepositional verb phrase to keep in touch in all but one occurrence in the corpus), as Excerpts 1-3 illustrate:

(1) would u touch my penis if u could?

(2) you ever touch yourself?

(3) did he touch u down there

Similarly, the word lick collocated saliently with the words for eight body parts, five of which were the names of sexual organs, four of them in their slang form: pussy, clit, ass, nipples, and cock. The other three body parts were not linked to sexual functions *per se* but acquired explicit sexual meaning through their collocational usage, namely: lick + legs; lick + fingers; lick + tongue. Concordance analysis of the remaining four key collocates of lick (lick + kiss/off/your/suck) revealed that they, too, served the online grooming strategy of sexual desensitisation, as Excerpts 4-6 illustrate.

(4) I wanna lick ur clit baby

(5) then I would spread your legs and lick the outside of you sweet pussy

(6) Would u let me lick u and kiss u all over every inch

In addition to illustrating the use of sexual desensitisation in online groomer discourse, the six examples considered above also show variation in terms of both im-/ex-plicitness

and (in)directness. Concerning im-/ex-plicitness, and as regards touch, Excerpt 3 avoids naming a sexual organ, replacing it instead with the euphemistic collocation down there; Excerpt 2 uses part-whole lexical relations whereby the whole (yourself, a non-sexual entity *per se*) is used to refer to a part (sexual organs). Excerpt 1, however, the sexual organ to be touched is explicitly stated. Similarly, and relating to lick, Excerpts 4 and 5 explicitly state their direct object, which is a sexual organ. In Excerpt 6, the groomer does not state the sexual organs that may be a direct object of TO LICK, choosing instead to refer explicitly and emphatically (all over… every inch) to his target as a whole (u). In other words, whereas all six examples contain sexually desensitising language, they are located at different points of a sexual implicitness – explicitness continuum.

As for (in)directness, considerable variation was also found. In Excerpt 4, for instance, the groomer uses a bald on record want statement to relay his sexual intention. Coupled with the use of sexual explicitness, this may be perceived to entail a high level of face threat, which may in turn account for the groomer's use of a term of endearment (baby). Through this positive politeness strategy, he may have tried to counter the impression of coercion behind his want statement. The other examples considered here all contain politeness strategies geared towards addressing the positive and/or negative face needs of the target. Negative politeness is evident in the use of conventionally indirect request formulae (would…) in Excerpts 1 and 6, as well as use of the modal auxiliary would in Excerpt 5, all of which suggest that the groomer is concerned with not being seen to impose on the target's freedom to decide. In Excerpt 5, the groomer uses positive politeness, framing his intention within an explicitly sexual compliment (you sweet pussy). Our analysis thus showed that sexual desensitisation was paired with sophisticated relational work geared towards addressing the target's positive and negative face needs. By being seemingly concerned about their interlocutors' face needs, online groomers also worked on their own face needs, seeking to come across as other-sensitive, rather than abusive, individuals.

4.2. Sexual Desensitisation in Online Grooming

As noted in Section 2, Black *et al.*'s (2015) study of online grooming language mapped the results of their LIWC analysis onto online grooming stages. Specifically, the study aligned the variable family with the risk assessment stage, along with the LIWC variables of pronouns (specifically the pronoun they), negative emotion (specifically, anxious) and home. Risk assessment is one of the aspects included within the broader online grooming process of compliance testing (see Figure 1). It is also, arguably, linked to the process of isolation, for some of the checks that online groomers will want to carry out when interacting online with their target concern the likelihood of their being able to gain exclusive online and/or offline access to him/her, whether physical (e.g. home) or affective (e.g. family and friends).

The LIWC2015 family variable contains 117 terms, ranging from `aunt*` to `wive*`. In our corpus, 205 terms were related to family, 60 of which occurred more than ten times. Table 3 shows the ten most dispersed of these (see also Table 3a for the top 50 list of family related terms, and their lexical dispersion values and frequency).

Rank	Word	DP_{Norm}	Frequency
209	mom	0.527	3297
341	dad	0.631	1810
467	parents	0.697	876
625	married	0.749	383
751	family	0.785	426
847	brother	0.811	284
852	wife	0.812	345
928	sister	0.826	293
995	daughter	0.388	286
1056	moms	0.847	218

Table 3: *Dispersion values for LIWC 'family' variable terms*

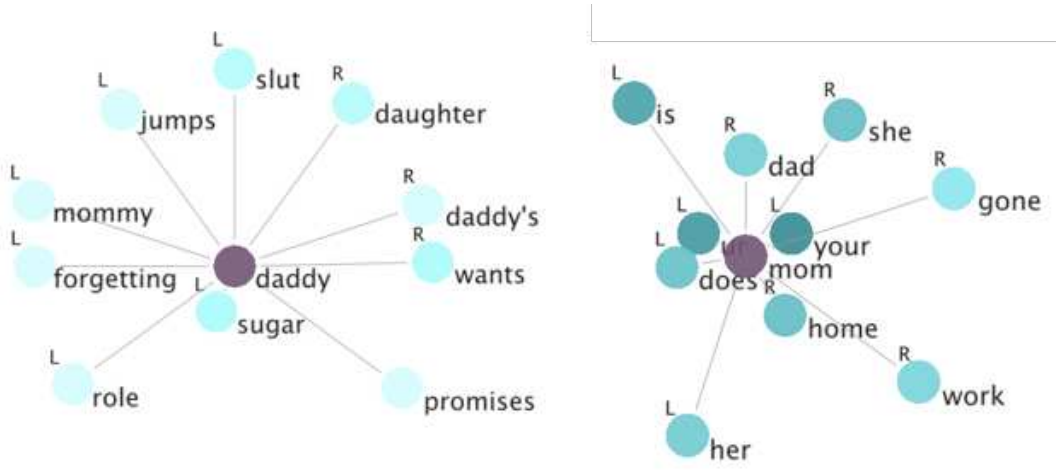

Figure 3: *Collocation graphs for 'daddy' and 'mom'*

Analysis of the top 15 collocates of these terms, and examination of their use in context through extended concordances, showed that family related words in our corpus were used to support more than one online grooming process. This can be seen through analysis of two of the family variable words with different dispersion values in the corpus, namely `daddy` (DP_{Norm} = 0.900) and 'mom' (DP_{Norm} = 0.527). Figure 3 shows their top 15 collocates, only two of which were sexually explicit: `daddy + sugar` and `daddy + slut`.

The concordance analysis of all the collocations in Figure 3 showed that they supported one or more of these online grooming processes: compliance testing, isolation and sexual gratification.

The concordance analysis of, in particular, the collocations mom + home/work/gone/does revealed that they were used by groomers repeatedly to gauge the likelihood of securing exclusive online and/or offline access to the target without his/her mother being around and/or aware of it. Consider excerpts 7 and 8:

(7) is your mom home with you now

(8) ur mom work every tues and weds?

The concordance analysis also showed that groomers recurrently sought to isolate their target emotionally from their parents, including by questioning their behaviour towards the target, as Excerpts 9 and 10 illustrate:

(9) do you forgive your mom for she has done to you and your brother?

(10) well i don't think your mom would like that if i got you pregnant

In both Excerpts, the groomer implicitly (9) or explicitly (10) evaluated the target's mother in negative terms, respectively as someone who had done something that may not be easily forgivable and as someone who would disapprove of the groomer-target's relationship, specifically of his 'g[etting] you [the target] pregnant'. By introducing this potential outcome (note the use of the conditional clause 'your mom would … if i...') of his intended relationship, the groomer also contributed to sexually desensitising his target. In this example, he did so implicitly, avoiding naming the activity – i.e., having sex – that would lead to the outcome (pregnancy).

Our concordance analysis clearly revealed the use of family terms for simultaneously advancing the online grooming processes of compliance testing, isolation and sexual gratification. This was especially the case regarding the term mom and, as with the illustrative examples in Section 4.1, groomers employed different levels of im-/ex-plicitness and (in)directness to realise these processes. As we saw in 10, and is also illustrated in Excerpt 11, sexual gratification was at times sought via use of implicit sexual desensitisation.

(11) so is your mom as cute as you?

Other times, however, family terms' key collocates were textually embedded within sexually explicit concordances, as can be seen in Excerpts 12 and 13.

(12) i'll be your sugar daddy

(13) go in the shower and shave your pussy baby. Daddy wants it nice and smooth

In Excerpt 12 the groomer self-allocated a role (sugar daddy) in his intended relationship with the target he was luring. The role was ostensibly sexual, and therefore sup-

ported the process of sexual gratification. In addition to explicit sexual lexis, the groomer's self-allocated role daddy also reframed his intended relationship in terms of power dynamics, whereby he – as the individual playing the dominant part – instructed his target, who was positioned as the dominated part. In Excerpt 13, the combination of sexual explicitness and absence of face mitigation (note the use of imperatives ('go in … shave') and a direct want statement (Daddy wants it…) may be perceived as coercive. Incidentally, these linguistic forms (imperatives and want statements) are not included in the list of overt persuasion / sexual extortion expressions in Schneevogt *et al.,* (2018). In Excerpt 13 the groomer may have once again tried to mitigate this through relational work, specifically the positive politeness strategy of using a term of endearment, baby, which presumes affective closeness. This may also apply to his self-positioning within the same turn as daddy in the next sentence, which may seek to affirm the close, indeed familial, bond he attached to their online 'relationship', and hence also to support the online grooming process of developing deceptive trust.

5. Conclusion

Research into online grooming language has primarily examined small data sets using qualitative (Thematic – and only recently Discourse) Analysis methods. In the few studies that have drawn upon larger data sets, the analysis has been quantitative and has tended to rely on software that operates at the decontextualized, single word level (LIWC2015). In view of this, this study sought to evaluate the contribution that an approach that provides 'a qualitative analysis of quantifiable patterns' (Marchi, 2010, p.164) – CADS – could make to advancing our understanding of the pressing social challenge of online grooming as a form of CSA. In doing so, we also wanted to determine how compatible a LIWC and CADS approach may be.

Our results show that both approaches were able to identify language that supported online grooming processes previously identified through Thematic Analysis, namely sexual gratification (through use of sexual terms), compliance testing and isolation (aligned with the family semantic field). However, it was only through a CADS methodology that we were able to capture the nuance behind contextual use of those processes.

Regarding sexual gratification, through CADS tools we were able to identify words included within the LIWC2015 dictionary for the sexual content variable that were not present in our corpus, for example those linked to sexually transmitted diseases. We were also able to identify a subset of words with sexual connotation in the corpus that were not included in the sexual content variable in LIWC2015 yet were both widely dispersed / frequent in our corpus and regularly used by groomers to desensitise their targets sexually. It was through calculation of their salient collocates, and manual analysis of concordances for these collocations, that we were able to identify how these words supported the sexual desensitisation strategy. The analysis also enabled our identifying important vari-

ation as regards im-/ex-plicitness and (in)directness levels. Moreover, strategic use of face-threat and politeness within the sexual desensitisation strategy was one of the findings that were only revealed through a CADS approach – not the use of LIWC – even though the software does contain grammatical and pragmatic variables that could, in theory, be clustered so as to generate that type of finding. The importance of being able to identify sexually implicit language patterns in online groomers' discourse cannot be underestimated. Similarly, becoming aware of levels of (in)directness in online groomers' discourse is also particularly important as it can help understand how groomers discursively navigate their choices along a 'persuasion-coercion' communicative continuum. Both im-/ex-plicitness and (in)directness are discourse features that – unlike decontextualized, single word analyses – a CADS approach can successfully examine.

As for risk assessment, this was linked in LIWC-based research to family terms. Our CADS approach showed that family terms, specifically 'mom' and 'daddy', were regularly used to advance, simultaneously, different grooming processes, namely sexual gratification (desensitisation and reframing), isolation (mental and physical) and deceptive trust development (relationship talk). This is not a surprising finding in Linguistics, where the multi-functionality of language has been long-established and is well-documented. However, it is an important finding within multi-disciplinary endeavours seeking to advance our knowledge of online groomers' communicative *modus operandi*, in which to date Linguistics studies have played a comparatively small but important part.

It is worth noting that LIWC was developed for the purpose of profiling psycholinguistic, rather than discourse, features. LIWC was certainly not developed to identify and profile online groomers' language in use, even though it has been deployed to that end. Thus, although LIWC should be credited with having significantly contributed to place language on the radar of multi-disciplinary research into online grooming, it may not be the most efficient software for understanding how online groomers communicatively entrap their targets. The lack of general language comparison scores, the fact that some of its variables are not transparent and its focus on decontextualized, single word analyses may make it a starting point for identifying some lexico-grammatical properties of online groomers' language.

Further work can now benefit from methods and tools in CADS, which can examine under-explored areas in previous research, including the differences between groomer-child and groomer-decoy interaction (Schneevogt *et al.,* 2018) as well as the complex ways in which explicit sexual desensitisation and face-work interact in online groomers' discourse. The analysis of corpora and discourse, as our results show, can shed light into the interwoven and non-linear essence of online grooming as a communicative entrapment network.

Finally, in this paper we have examined linguistic patterns for some of the processes and strategies that, according to the literature, characterise online grooming. Having es-

tablished that CADS can make a positive contribution, future work on online grooming should continue to apply this approach to a range of pressing issues, including determining how extra-linguistic factors, such as duration of grooming, may reflect and shape different 'groomer' communicative profiles beyond compliment use. This line of work, geared towards developing a nuanced understanding of online groomers' communicative *modus operandi,* is crucial when it comes to working with law enforcement and other stakeholders (e.g. educationalists, social workers) on developing targeted interventions for preventing CSA online.

Acknowledgements

The research reported here is part of the Swansea University-led, multi-disciplinary project Developing Resilience against Online Grooming (DRaOG) – (for details, please contact: developingresilienceagainstOG@gmail.com). We are also grateful to Dr Paiement, Dr Di Cristofaro, Mr Morgan, Dr Reitmaier and Mr Martinez-Granado for their assistance regarding data extraction.

Funding

We would like to thank CHERISH-DE (http://cherish-de.uk/) for funding received in relation to work undertaken within the remit of this project.

Competing interests

The authors have no competing interests to declare.

Appendices

Rank	word	DP_{Norm}	frequency
210	sex	0.529	3715
248	sexy	0.597	2899
427	naked	0.673	1580
457	pussy	0.690	2507
459	cum	0.691	2314
586	horny	0.732	1190
660	fuck	0.759	1360
1440	dick	0.777	1440
787	pregnant	0.796	276
788	cock	0.797	1457
808	virgin	0.802	421

933	fucking	0.827	459
948	sexual	0.830	266
975	clit	0.834	324
1043	boobs	0.846	263
1133	porn	0.858	553
1147	nude	0.859	551
1194	penis	0.865	465
1196	sexually	0.866	150
1217	fucked	0.868	217
1243	tits	0.871	406
1263	condom	0.874	254
1277	orgasm	0.876	247
1306	masturbate	0.879	260
1318	condoms	0.880	204
1327	gay	0.882	235
1379	breast	0.888	143
1435	breasts	0.893	269
1591	cumming	0.906	111
1725	screwed	0.916	43
1784	blowjob	0.919	151
1844	vagina	0.923	175
1856	passion	0.923	50
1867	erotic	0.924	37
1871	perv	0.924	72
1873	nipple	0.925	47
1877	sexiest	0.925	33
1930	lover	0.927	116
2031	rape	0.932	91
2065	bi	0.932	73
2266	screw	0.941	41
2347	pervert	0.943	42
2364	masturbating	0.944	65
2407	pubic	0.945	25
2646	slut	0.951	264
2652	lovers	0.952	35

2657	cums	0.952	41
2694	sexier	0.953	29
2776	foreplay	0.954	47
2856	perverted	0.956	32

***Table 2a**: Top 50 sexual words in the Perverted Justice corpus*

Rank	word	DP_{Norm}	frequency
209	mom	0.527	3297
341	dad	0.631	1810
467	parents	0.697	876
528	baby	0.715	6430
625	married	0.749	383
751	family	0.785	426
787	pregnant	0.796	276
847	brother	0.811	284
852	wife	0.812	345
928	sister	0.826	293
995	daughter	0.838	286
1056	moms	0.847	218
1061	son	0.848	231
1167	marry	0.863	194
1175	mother	0.864	159
1203	dads	0.866	157
1226	brothers	0.869	131
1260	grandma	0.874	180
1266	sis	0.875	356
1310	sisters	0.880	145
1440	father	0.893	125
1517	daddy	0.900	167
1605	babies	0.908	52
1717	uncle	0.915	109
1766	divorced	0.918	80
1864	aunt	0.924	139
1908	bro	0.927	123
2068	wedding	0.933	42

2151	niece	0.936	63
2166	cousins	0.937	46
2193	husband	0.938	48
2314	grandpa	0.942	71
2344	marriage	0.943	35
2349	mom's	0.943	67
2507	cousin	0.948	103
2535	divorce	0.949	34
2568	nephew	0.950	39
2666	pa	0.952	34
2857	sons	0.956	24
2887	dad's	0.957	40
2958	parent	0.958	56
3077	mommy	0.960	37
3080	ma	0.960	82
3170	daughters	0.962	38
3669	nieces	0.969	20
3995	aunts	0.972	38
4037	grandparents	0.973	27
5063	folks	0.980	45

***Table 3a**: Top 50 family words in the Perverted Justice corpus*

Number	Position	Collocate	Dice-coefficient
1	R	yourself	0.079
2		touch	0.057
3	L	kiss	0.042
4	R	pussy	0.033
5	R	penis	0.031
6	R	vagina	0.028
7	R	body	0.028
8	R	tits	0.028
9	L	ever	0.026
10	L	did	0.026
11	L	keep	0.026
12	R	self	0.024

13	R	breasts	0.024
14	R	lips	0.023
15	L	anxious	0.02

Table 4: *Dice-coefficient and position of top 15 collocates of 'touch'*

Number	Position	Collocate	Dice-coefficient
1	R	pussy	0.160
2	R	suck	0.085
3	R	nipples	0.064
4	R	clit	0.046
5		lick	0.046
6	L	legs	0.038
7	L	kiss	0.032
8	R	finger	0.032
9	M	cock	0.029
10	R	your	0.028
11	R	tongue	0.028
12	R	ass	0.027
13	R	off	0.026
14	R	ur	0.024
15	R	over	0.024

Table 5: *Dice-coefficient and position of top 15 collocates of 'lick'*

References

Baker, P., & Levon, E. (2015). Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity. *Discourse & Communication, 9*(2), 221-236. doi:10.1177/1750481314568542

Baron, A. (2018). *VARD – Variant Detector (Version 2.5.4)*. [Computer Software]. Retrieved from: http://ucrel.lancs.ac.uk/vard/about/

Bentley, H., Burrows, A., Clarke, L., Gillgan, A., Glen, J. Hafizi,M., Letendrie, F., Miller, P. O'Hagan, O. Patel, P., Peppiate, J., Stanley, K. Starr, E., Vasco, N., & Walker, J. (2018) *How safe are our children? The most comprehensive overview of child protection in the UK*. London: NSPCC. Retrieved from https://learning.nspcc.org.uk/media/1067/how-safe-are-our-children-2018.pdf

Black, P., Wollis, M., Woodworth, M., & Hancock, J.T. (2015). A linguistic analysis of grooming strategies of online child sex offenders: Implications for our understanding of predatory sexual behavior in an increasingly computer-mediated world. *Child Abuse & Neglect, 44,* 140-149. doi:10.1016/j.chiabu.2014.12.004

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics, 20*(2), 139-173. doi:10.1075/ijcl.20.2.01bre

Brookes, G., & McEnery, T. (2018) The utility of topic modelling for discourse studies: A critical evaluation. *Discourse Studies*, 1-19. doi:10.1177/1461445618814032

Craven, S., Brown, S., & Gilchrist, E. (2006). Sexual grooming of children: Review of literature and theoretical considerations. *Journal of Sexual Aggression, 12,* 287-299. doi:10.1080/13552600601069414

Chiang, E., & Grant, T. (2017). Online grooming: moves and strategies. *Language and Law/Linguagem e Direito, 4*(1), 103-141.

Chiang, E., & Grant, T. (2018). Deceptive identity performance: Offender moves and multiple identities in online child abuse conversations. *Applied Linguistics,* 1-25. doi:10.1093/applin/amy007

De Choudhury, M., Counts, S., & Horvitz, E. (2013). Social Media as a Measurement Tool of Depression in Populations. *Proceedings of WebSci'13, Paris, France*. 47-56. https://doi.org/10.1145/2464464.2464480

Drouin, M., Boyd, R., Hancock, J.T., & James, A. (2017). Linguistic analysis of chat transcripts from child predator undercover sex stings. *The Journal of Forensic Psychiatry & Psychology, 28*(4), 437-457. doi:10.1080/14789949.2017.1291707

Egan, V., Hoskinson, J., & Shewan, D. (2011). Perverted Justice: A Content Analysis of the Language used by Offenders Detected Attempting to Solicit Children for Sex. In R. M. Clark (Ed.), *Antisocial Behaviour: Causes, Correlations and Treatments* (pp. 119-133). New York: Nova Science Publishers.

Gabrielatos, C. (2018). Keyness Analysis: nature, metrics and techniques. In C. Taylor & A. Marchi (Eds.), *Corpus Approaches to Discourse: A Critical Review* (pp. 225-258). London: Routledge.

Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International journal of corpus linguistics, 13*(4), 403-437. doi:10.1075/ijcl.13.4.02gri

Gries, S.T. (2010). Useful statistics for corpus linguistics. In A. Sánchez & M. Almeda (Ed.), *A Mosaic of Corpus Linguistics: Selected Approaches* (pp. 269-291). Frankfurt am Main: Peter Lang.

Hardie, A. (2012). CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics, 17*(3), 380–409. doi:10.1075/ijcl.17.3.04har

Hamilton-Giachritsis, C., Hanson, E., Whittle, H., & Beech, A. (2017). *Impact of online and offline child sexual abuse: "Everyone deserves to be happy and safe"*. NSPCC. Retrieved from https://learning.nspcc.org.uk/research-resources/2017/impact-online-offline-child-sexual-abuse/

Herring, S. (2013). Discourse in Web 2.0: Familiar, reconfigured, and emergent. In D. Tannen & A. M. Tester (Eds.), *Georgetown University Round Table on Language and Linguistics 2011: Discourse 2.0: Language and New Media* (pp. 1-25). Washington DC: Georgetown University Press.

Himelboim, I., Gleave, E., & Smith, M. (2009) Discussion catalysts in online political discussions: Content importers and conversation starters. *Journal of Computer Mediated Communication, 14*(4), 771-789. doi:10.1111/j.1083-6101.2009.01470.x

Katz, C. (2013). Internet-related child sexual abuse: What children tell us in their testimonies. *Children and Youth Services Review, 35,* 1536–1542. doi:10.1016/j.childyouth.2013.06.006

Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun Use Reflects Standings in Social Hierarchies. *Journal of Language and Social Psychology, 33*(2), 125-143. doi:10.1177/0261927X13502654

Kloess, J.A., Hamilton-Giachritsis, C.E., & Beech, A.R. (2017). Offense Processes of Online Sexual Grooming and Abuse of Children Via Internet Communication Platforms. *Sexual Abuse 31(1),* 1-24. doi:10.1177/1079063217720927

Lorenzo-Dus, N., & Di Cristofaro, M. (2016) #Living/minimum wage: Influential citizen talk in twitter. *Discourse, Context & Media, 13.* 40–50. https://doi.org/10.1016/j.dcm.2016.02.003

Lorenzo-Dus, N., Izura, C., & Pérez-Tattam, R. (2016). Understanding grooming discourse in computer-mediated environments. *Discourse, Context & Media, 12,* 40-50. doi:10.1016/j.dcm.2016.02.004

Lorenzo-Dus, N., & Izura, C. (2017). "Cause ur special": Understanding trust and complimenting behaviour in online grooming discourse. *Journal of Pragmatics, 112,* 68-82. doi:10.1016/j.pragma.2017.01.004

Marchi, A. (2010) The 'moral *in* the story': a diachronic investigation of lexicalised morality in the UK press. *Corpora, 5,* 161-189. doi:10.3366/cor.2010.0104

Marchi, A., & Taylor, C. (2018) Introduction: Partiality and reflexivity. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse. A critical review* (pp.1-15). London: Routledge.

O'Connell, R. (2003). *A typology of child cybersexploitation and online grooming practices*. Preston, England: University of Central Lancashire. Retrieved from: http://image.guardian.co.uk/sys-files/Society/documents/2003/07/17/Groomingreport.pdf

O'Dea, B., Larsen, M. E., Batterham, P. J., Calear, A.L., & Christensen, H. (2017). A linguistic analysis of suicide-related Twitter posts. *Crisis, 38*(5), 319-329. doi:10.1027/0227-5910/a000443

Park, M., Cha, C., & Cha, M. (2012). Depressive moods of users portrayed in Twitter. In Proceedings of ACM SIGKDD Workshop on Healthcare Informatics (HI-KDD).

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015a). *The development and psychometric properties of LIWC2015.* Austin: University of Texas at Austin, TX. Retrieved from https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf

Pennebaker, J.W., Booth, R.J., Boyd, R.L., & Francis, M. E. (2015b). *Linguistic Inquiry and Word Count: LIWC2015: Operator's Manual.* Austin: University of Texas at Austin, TX. Retrieved from https://s3-us-west-2.amazonaws.com/downloads.liwc.net/LIWC2015_OperatorManual.pdf

Quayle, E., & Newman, E. (2017). An exploratory study of public reports to investigate patterns and themes of requests for sexual images of minors online. *Crime Science, 5*(2), 1-12. doi:10.1186/s40163-016-0050-0

Quercia, D., Ellis, J., Capra, L., & Crowcroft, J. (2011) In the mood for being influential on twitter. In Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third

International Conference on Social Computing (SocialCom), 307-314. Piscataway: IEEE.

Schneevogt, D., Chiang, E., & Grant, T. (2018). Do Perverted Justice chat logs contain examples of overt persuasion and sexual extortion? A research note responding to Chiang and Grant (2017, 2018). *Language and Law/Linguagem e Direito, 5*(1), 97-102.

Tausczik, Y.R., & Pennebaker, J.W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24-54. doi:10.1177/0261927X09351676

van Dijk, T. A. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, N.J: Lawrence Erlbaum.

van Gijn-Grosvenor, E. L., & Lamb, M.E. (2015). Behavioural differences between online sexual groomers approaching boys and girls. J*ournal of Child Sexual Abuse. 25*(5), 577-596. doi:10.1080/10538712.2016.1189473

Vergani, M., & Bliuc, A. (2018). The language of new terrorism: Differences in psychological dimensions of communication in Dabiq and Inspire. *Journal of Language and Social Psychology, 37*(5), 523-540. doi:10.1177/0261927X17751011

Whittle, H., Hamilton-Giachritsis, C.E., & Beech, A. (2014). "Under his spell": Victims' perspectives of being groomed online. *Social Sciences, 3,* 404-426. doi:10.3390/socsci3030404

Williams, R., Elliott, I.A., & Beech, A. (2013). Identifying sexual grooming themes used by Internet sex offenders. *Deviant Behavior, 34*(2), 135-152. doi:10.1080/01639625.2012.707550

Wilson, M.L., Ali, S. A., & Valstar, M.F. (2014). Finding information about mental health in microblogging platforms: a case study of depression. IIiX '14 Proceedings of the 5th Information Interaction in Context Symposium, 8-17. Regensburg, Germany. https://doi.org/10.1145/2637002.2637006

Winters, G. M., Kaylor, L.E., & Jeglic, E.L. (2017). Sexual offenders contacting children online: an examination of transcripts of sexual grooming. *Journal of Sexual Aggression, 23*(1), 62-76. doi:10.1080/13552600.2016.1271146